Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Gudbjartsson DF, Helgason A, Jonsson H, et al. Spread of SARS-CoV-2 in the Icelandic population. N Engl J Med. DOI: 10.1056/NEJMoa2006100

TABLE OF CONTENT

LIST OF INVESTIGATORS

Investigator	Affiliation
Daniel F Gudbjartsson	deCODE genetics/Amgen, Inc., Reykjavik, Iceland. School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland.
Agnar Helgason	deCODE genetics/Amgen, Inc., Reykjavik, Iceland. Department of Anthropology, University of Iceland, Reykjavik, Iceland.
Hakon Jonsson	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Olafur T Magnusson	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Pall Melsted	deCODE genetics/Amgen, Inc., Reykjavik, Iceland. School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland.
Gudmundur L Norddahl	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Jona Saemundsdottir	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Asgeir Sigurdsson	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Patrick Sulem	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Arna B Agustsdottir	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Berglind Eiriksdottir	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Run Fridriksdottir	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Elisabet E Gardarsdottir	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Gudmundur Georgsson	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Olafia S Gretarsdottir	Department of Clinical Microbiology, Landspitali - The National University Hospital, Reykjavik, Iceland.
Kjartan R Guðmundsson	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Thora R Gunnarsdottir	Department of Clinical Microbiology, Landspitali - The National University Hospital, Reykjavik, Iceland.
Arnaldur Gylfason	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Hilma Holm	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Brynjar O Jensson	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Aslaug Jonasdottir	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Frosti Jonsson	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Kamilla S Josefsdottir	Directorate of Health, Reykjavik, Iceland.
Thordur Kristjansson	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Droplaug N Magnusdottir	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Louise le Roux	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Gudrun Sigmundsdottir	Directorate of Health, Reykjavik, Iceland
Gardar Sveinbjornsson	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Kristin E Sveinsdottir	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Maney Sveinsdottir	Department of Clinical Microbiology, Landspitali - The National University Hospital, Reykjavik, Iceland.
Emil A Thorarensen	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Bjarni Thorbjornsson	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.
Arthur Löve	Department of Clinical Microbiology, Landspitali - The National University Hospital, Reykjavik, Iceland. Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland.
Gisli Masson	deCODE genetics/Amgen, Inc., Reykjavik, Iceland.

Investigator	Affiliation
Ingileif Jonsdottir	deCODE genetics/Amgen, Inc., Reykjavik, Iceland. Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland.
Alma D. Möller	Directorate of Health, Reykjavik, Iceland.
Thorolfur Gudnason	Directorate of Health, Reykjavik, Iceland.
Karl G Kristinsson	Department of Clinical Microbiology, Landspitali - The National University Hospital, Reykjavik, Iceland. Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. BioMedical Centre of the University of Iceland, Reykjavik, Iceland.
Unnur Thorsteinsdottir	deCODE genetics/Amgen, Inc., Reykjavik, Iceland. Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland.
Kari Stefansson	deCODE genetics/Amgen, Inc., Reykjavik, Iceland. Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland.

SUPPLEMENTARY METHODS

SARS-COV-2 TESTING

SARS-CoV-2 infected individuals were identified in Iceland through targeted testing and population screening. A flowchart outlining the study is shown in **Figure 1**.

The targeted testing began on January 31 2020 and focused on individuals deemed at high-risk of being infected by SARS-CoV-2. These were mainly symptomatic (cough, fever, aches, and shortness of breath) individuals returning to Iceland from countries/regions classified by the health authorities as high-risk, or who had been in contact with infected individuals. Individuals that met the criteria for targeted testing were encouraged by the health authorities to contact the health care system for testing. They could call a specific phone number, connect through a specific website or contact their primary health care provider. A contagion tracker team tracked down contacts of infected individuals and arranged testing if they had symptoms. Sample for testing were taken in most local primary health care clinics and hospitals. In addition, a number of individuals with symptoms were tested at the international airport on arrival from high-risk areas.

As of March 19, all travel outside Iceland was designated high-risk. As of March 31, 9,199 individuals had been targeted for testing.

The population screening for SARS-CoV-2 was initiated by deCODE on March 13. The screening was open to all Icelandic residents who were symptom free or with mild symptoms of common cold that is highly prevalent in Iceland at this time of the year. The registration for the test was online and during sample collection information on recent travels, contacts with infected individuals, and symptoms compatible with COVID-19 were registered. The sample collection was performed in Reykjavik, the capital of Iceland. On April 1, after twenty days of population screening, 10,797 Icelanders had been screened for SARS-CoV-2.

To evaluate the sampling method of the population screening, we also invited 6,782 randomly chosen Icelanders between ages 20 and 70 to participate through a phone text message sent on March 31 and April 1. Of these, 2,283 had participated by April 4 (33.7%). Of the invited, 41.2% were males and of the participants 37.8% were males.

Those who tested positive for SARS-CoV-2 through any means were required to self-isolate, and those who had been in contact with them, to self-quarantine, for 2 weeks. In addition to isolating positives and quarantining those at high risk of infection, on March 16 the Icelandic authorities initiated a ban on gathering of over 100 people and stated that a social distancing of at least 2 meters should be maintained. On March 24, the gatherings were restricted to 20 people. To protect the elderly and other groups that are at greater risk of serious illness should they develop COVID-19 health authorities have promoted self-isolation and banned visits to nursing homes and hospitals. Although universities and colleges have been closed since March 16, daycares and elementary schools have been open. For SARS-CoV-2 testing, it was recommended to take both nasopharyngeal and oropharyngeal samples. RNA from all samples was isolated within 24 hours.

TRACKING OF SARS-COV-2 INFECTIONS

All individuals who tested positive for SARS-CoV-2 were contacted by phone by a contagion tracking team designated by the authorities. They were instructed to enter isolation at home. They were asked about their symptoms and onset, recent travels and previous contacts with infected individuals. They were also asked to identify everyone whom they had been in contact with 24 hours before noticing their first symptom and for how long they interacted with each individual and how intimate the interaction was. All registered contacts were contacted by phone, requested to go into 2 weeks quarantine and asked about symptoms.

Those with symptoms and those who developed them in quarantine were tested for SARS-CoV-2.

Isolation places more severe restrictions on the individual than quarantine. The Icelandic surgeon general provides detailed instructions about quarantine and isolation on the webpages https://www.covid.is/categories/how-does-quarantine-work and https://www.covid.is/categories/how-does-isolation-work. Quarantine last for fourteen days. Isolation ends 10 days after fever subsides or when the individual tests negative for presence of the virus.

RNA EXTRACTION

Viral RNA samples were extracted either at the Department of Clinical Microbiology laboratory at Landspitali - the National University Hospital of Iceland (LUH) or at deCODE. Both extraction methods are based on an automated magnetic bead-purification procedure, which includes cell lysis and Proteinase K treatment. RNA from samples at LUH were extracted (32 samples per 60 min run) using the MagNA Pure LC 2.0 or MagNA Pure Compact instruments from Roche LifeScience, with 200/100 µL input/output volume(s), respectively. Samples at deCODE were extracted from swabs (96 samples per 70 min run) using the Chemagic Viral RNA kit on the Chemagic360 instrument from Perkin Elmer, with 300/100 µL input/output volume(s), respectively. Each step in the workflow was monitored using an in-house LIMS (VirLab) with 2D barcoding (Greiner, 300 µL tubes) of all extracted samples.

TESTING OF SAMPLES FOR SARS-COV-2

Testing for SARS-CoV-2 was performed either at LUH or deCODE using similar quantitative real-time PCR (qRT-PCR) methods. The assay at LUH is based on the WHO recommended

screening method (https://www.who.int/docs/default-source/coronaviruse/wuhan-virus-assay-v1991527e5122341d99287a1b17c111902.pdf), which involves a single probe pan-screening assay for betacoronaviruses, followed by confirmatory measurements for all positive samples using an nCoV-2019 specific assay. The broad betacoronavirus assay is based on probes for a conserved region of the E-gene, whereas confirmatory testing assays were done using either nCoV-2019 specific probes for the RdRp gene or the TaqMan™ Fast Virus 1-step Master Mix, 2019-nCoV Assay kits v1 from Thermo Fisher (three probes, see below for details). All labelled probes and primers for the E-and -RdRP genes were from TAG (Copenhagen, Denmark). Superscript™ III One-Step RT-PCR assay mix with Platinum™ *Taq* DNA polymerase was from ThermoFisher. 2019 E gene control and SARS-CoV Frankfurt 1 positive controls were obtained from EVAg (https://www.european-virus-archive.com/bundle/diagnostics-controls-wuhan-coronavirus-2019-2019-ncov). Each assay was done in a 25 µL total sample volume with FAM™ dye labelled probes in addition to VIC™ dye labelled probes for human RNase P as internal control. Plates (96 well) were scanned in an AB-7500 Fast real-time PCR thermocycler for 40 cycles of amplification following the manufacturer´s instructions (ThermoFisher). Samples in the E-gene screening assay with C_t <35 were considered strong positive and went for confirmatory testing using RdRp, whereas samples with C_t values between 35-37 were considered weak positive and were confirmed using the TaqMan™ Fast Virus method. Samples with C_t values from 37-40 were classified as inconclusive and were tested again to confirm their status. The sensitivity of the WHO assays are 5.2 RNA copies/reaction for the non-specific E-gene and 3.8 RNA copies/reaction for the 2019- nCoV specific confirmatory RdRp gene. Specificity of the WHO recommended assays were assessed against a number of known viruses, including alphacoronaviruses, non-asian strains of betacoronaviruses, influenza and MERS. No cross-reactivity was observed.

deCODE used exclusively the three probe TaqMan™ Fast Virus 1-step Master Mix, 2019-nCoV v1 assay described above. Both the assay kit and 2019-nCov control kits were obtained from Thermo Fisher. Assay mix A, B and C were prepared containing FAM™ dye labelled probes for the SARS-CoV-2 specific genes ORF1ab, S-protein and N-protein, respectively. In addition, each assay mix contained VIC™ dye labelled probes for human RNase P as internal control. Samples from 96-well RNA sample plate(s) were dispensed into three wells each in a 384 plate layout, in addition to three negative (no template) and three positive controls. Assay mix was added in a total reaction volume of 12.5 µL per sample. All sample aliquoting and mixing at deCODE was performed with an automated Hamilton STARlet 8-channel liquid handler and the assay plates were scanned in an ABI 7900 HT RT-PCR system following manufacturer´s instructions with a total of 40 cycles of amplification. Samples with FAM™ dye C_t values <37 in at least two of three assays were classified as positive. Samples with FAM™ dye C_t values between 37 and 40 were classified as inconclusive and their testing repeated. If repeated testing gave the same result with at least two probes the sample was classified as positive. If repeated testing gave positive results for only one probe the test was considered inconclusive and a new sample from the subject was requested. The frequency of inconclusive results was 0.04% (data not shown). Samples with undetected FAM™ dye C_t values or values equal to 40 in all three assays were classified as negative if the human RNaseP assay was positive (VIC™ dye C_t <40). The sensitivity of the assay was evaluated by serial dilution of the positive control and was estimated at 6 copies per reaction (data not shown). Validation of the RNA extraction and the qRT-PCR method(s) at deCODE was performed using 124 samples that had previously tested positive (n=104) or negative (n=20) with the qRT-PCR assay at LUH. All of the negative samples tested negative at deCODE and 102 of the 104 positive tested at LUH were also positive at

deCODE. Two samples that tested positive at LUH were negative at deCODE. Upon subsequent sequencing (see below) viral genome could not be detected in these two samples, probably because very few viral particles were present. Samples from 643 individuals that tested positive using either the deCODE or the LUH qPCR assays were also submitted for viral genome sequencing (see below). Viral RNA (cDNA) from six samples (0.9 %) yielded no sequence data mapping to the viral reference genome. The success of generating sequencing libraries with good coverage is highly dependent on the amount of viral RNA in the samples as assessed by the C_t values from the qRT-PCR assays. Figure S2 shows the relationship between measured C_t values and the consensus coverage of the sequenced samples. These data show that the qRT-PCR assay is more sensitive in detecting viral RNA than the amplicon sequencing method.

SAMPLE PREPARATION FOR SEQUENCING

Reverse transcription (RT) and multiplex PCR was performed based on information provided by the Artic Network initiative (https://artic.network/) to generate cDNA. In short, extracted viral RNA was pre-incubated at 65 °C for 5 min in the presence of random hexamers (2.5 µM) and dNTP´s (500 µM). Sample cooling on ice was then followed by RT using SuperScript IV (ThermoFisher) in the presence of DTT (5 mM) and RNaseOUT inhibitor (ThermoFisher) for 10 min at 42°C, followed by 10 min at 70 °C. Multiplex PCR of the resulting SARS-CoV-2 cDNA was performed using a tiling scheme of primers, designed to generate overlapping amplicons of approximately 800 bp (Table S1). The primers were generously provided by Dr. David Stoddard at Oxford Nanopore Technologies. Two PCR reactions were done for each sample using primer pools A and B, respectively (Table S1). PCR amplification was done using the Q5® Hot Start High-Fidelity polymerase (New England Biolabs) with primers at 1 µM concentration. The reactions were performed in a MJR thermal cycler with a heated lid at 105 °C, using 35 cycles

of denaturation (15 sec at 98 °C) and annealing/extension (5 min at 65 °C). The resulting PCR amplicons were purified using Ampure XP magnetic beads (Beckman Coulter) and quantified using the Quant-iT™ PicoGreen dsDNA assay kit (Thermo Fisher). Amplified samples (20-500 ng) were randomly sheared using focused acoustics in 96-well AFA-TUBE-TPX plates (Covaris Inc.) on the Covaris LE220plus instrument with the following settings: Sample volume, 50 µL; temperature, 10 °C; peak incident power, 200W; duty factor, 25%; cycles per burst, 50; time, 350 sec. Sequencing libraries were prepared in the 96 well Covaris plates, using the NEBNext® Ultra II kit (New England Biolabs) following the manufacturer´s instructions. In short, end repair and A-tailing was performed in a combined reaction per sample (plate) for 30 min at 20 °C, followed by thermal enzyme inactivation at 65 °C for 30 min. Adaptor ligation was done using the NEBNext® ligation master mix plus enhancer and the TruSeq unique dual indexed IDT adaptors (Illumina, **Table S2**). Ligation reactions were incubated for 15 min at 20 °C. Ligated sequencing libraries were purified on a Hamilton STAR NGS liquid handler, using two rounds of magnetic SPRI bead purification (0.7X volume).

ILLUMINA SEQUENCING

Sequencing libraries were pooled (24-36 samples/pool) and quantified using the Qubit dsDNA assay (ThermoFisher). Samples were diluted appropriately and denatured to a final loading concentration of 10 pM. All samples were sequenced on Illumina MiSeq sequencers using 300-cycle MiSeq v2 reagent kits (Illumina). Each pool was sequenced using dual indexed paired-end sequencing of 150*8*8*150 bp cycles of data acquisition and imaging with a run time of approximately 24 hrs. Basecalling was done in real time using MCS v3.1 and FASTQ files were generated using MiSeq Reporter. At least 15M PF reads (>4.5Gb) with base qualities of >Q30 for at least 90% of bases were collected for each run.

SEQUENCING DATA ANALYSIS

Amplicon sequences were aligned to the reference genome of the SARS-CoV-2 (NC_045512.2)[1] using bwa mem[2], possible PCR duplicates were marked with markduplicates from Picard tools[3] and reads with less than 50 bases aligned were omitted from the alignment. The resulting aligned filtered reads were used for variant calling with bcftools[2]. For consensus sequence generation only variants reported as homozygous were used. In regions targeted with primers we allowed variants to have allele frequency below one in individual. The consensus sequence was masked with ambiguous nucleotides (N) at positions if the depth of coverage was strictly less than 5 reads after restricting to bases of quality 20 or higher. Consensus sequences with more than 10,000 ambiguous nucleotides were discarded from analysis. The SARS-COV-2 consensus sequences with less than 10,000 ambiguous nucleotides were uploaded to GISAID (https://www.gisaid.org), with virus names hCov/Iceland/1/2020-hCov/Iceland/604/2020, and EPI accession numbers, EPI_ISL_417481, EPI_ISL_417535-EPI_ISL_417876, and EPI_ISL_424367-EPI_ISL_424624. The mutations in Table S3 were used to define haplogroups/clades.

In addition to calling consensus sequence of the samples independently, we jointly called variants across the samples using a modified version of Graphtyper.[4] In this modified version we excluded the termini of paired end reads overlapping a primer region. Further we supplied the following flags to Graphtyper:

--no_filter_on_read_bias,--no_filter_on_strand_bias --no_filter_on_coverage --impurity_threshold=1.0 --primer_bedpe={primer_file} --is_only_cigar_discovery --genotype_aln_min_support_ratio=0.30 --genotype_aln_min_support=5 --is_discovery_only_for_paired_reads --no_filter_on_begin_pos

In this joint calling, we restricted to genotypes with coverage of 5 or more. We then define a carrier matrix, where the cell in matrix was 1 if the fraction of reads supporting the alternative allele (AB) was greater than 85% (carrier), and 0 if the AB was less than 5% (non-carrier). With this carrier matrix, we manually curated list of mutations in in Table S3 to define haplotypes. We then assigned a haplotype per sample, by aggregating the haplotype informative mutations of the sample and constructing a consensus haplotype.

Variants were annotated using Variant Effect Predictor (VEP) version 99.2[5] in custom mode. As a gene map we used Uniprot mature gene track downloaded from UCSC[6]. A single VEP annotation was assigned to each sequence variant according to our previously described scheme (Sveinbjornsson et al 2015) using gorpipe scripts[7].

For network analysis of haplogroups a median-joining network[8] of SARS-CoV-2 sequences was generated using data from our sequencing effort in Iceland and from GISAID available on March 22 (Table S4). Only sequences with start positions <=200 and stop positions >=29750 were included in the analysis. For the GISAID sequences, only those with <=1% missing nucleotides were used, whereas for the Icelandic sequences a more permissive threshold of <=5% was imposed. To reduce noise in the network, an imputation step was implemented for sequences with missing nucleotides at sites where other sequences varied, whereby the missing nucleotide was imputed to the consensus variant for the clade it was assigned to, based on non-missing sites.

Contact tracking information was obtained from the Chief Epidemiologist, Directorate of Health, for 1317 confirmed cases. The information includes travel abroad, confirmed transmissions, contact with other confirmed cases as well as demographic information. The

study was approved by the National Bioethics Committee of Iceland (Approval no. VSN-20-070).

SUPPLEMENTARY FIGURES

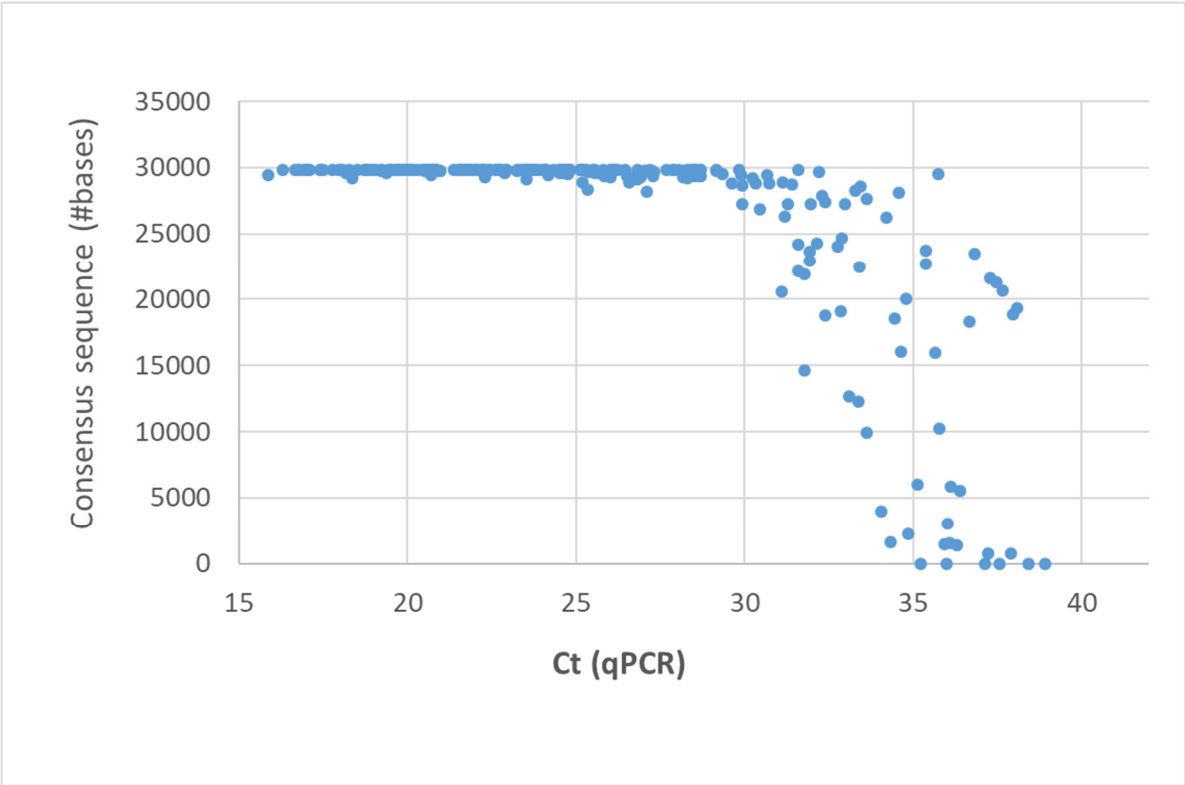
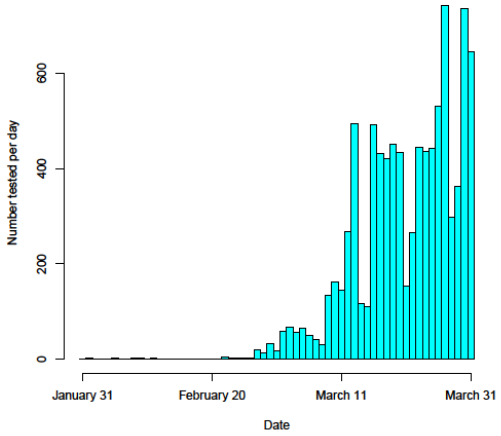
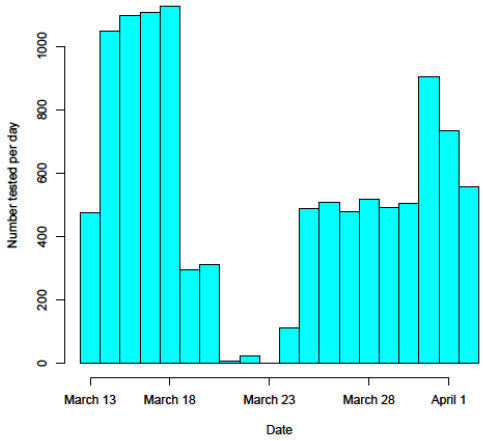


Figure S1 Coverage of sequenced viral genomes shown as size of the consensus sequence as a function of measured Ct values from the qRT-PCR screening assays. A total of 643 samples were sequenced, thereof 581 and 605 samples with coverage for at least 90% (27Kb) and 67% (20 Kb) of the genome, respectively. Six samples yielded no coverage.

A



B



C

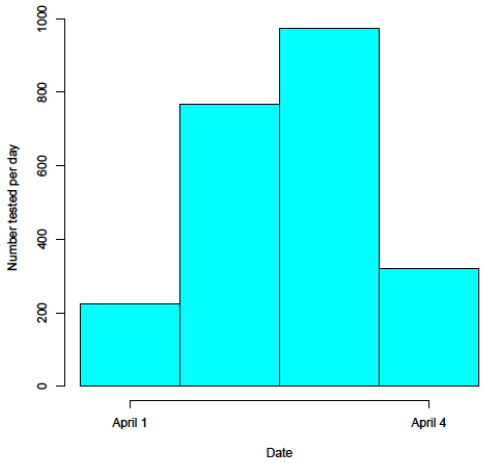
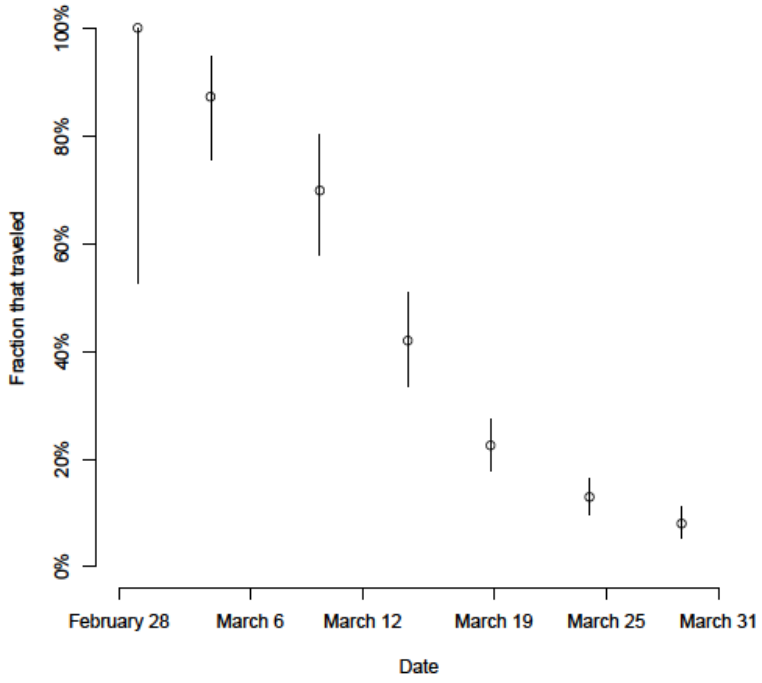


Figure S2 The number of individuals tested per day (A) in the targeted testing, (B) the open invitation part of the population screening, and (C) the random sample from the population screening.

A



B

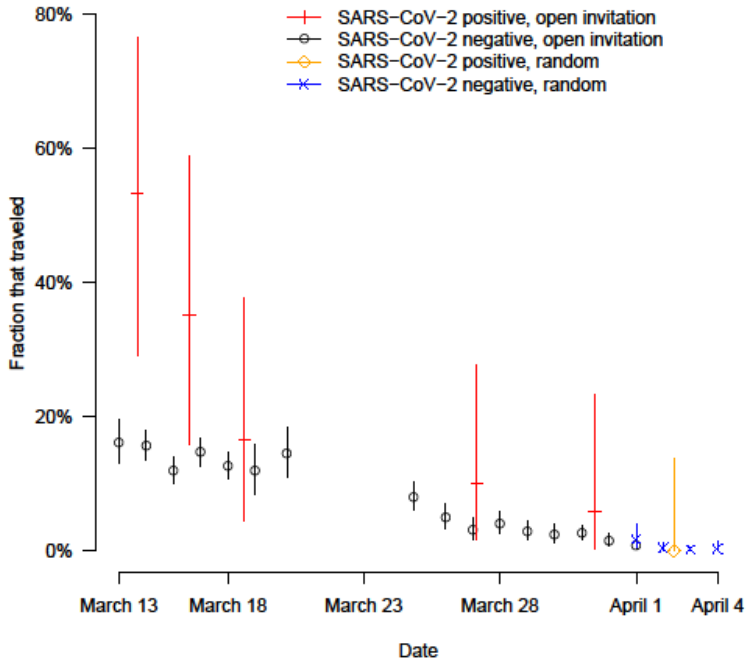


Figure S3 Fraction of participants in (A) the targeted testing and (B) population screening that had recently traveled outside Iceland.

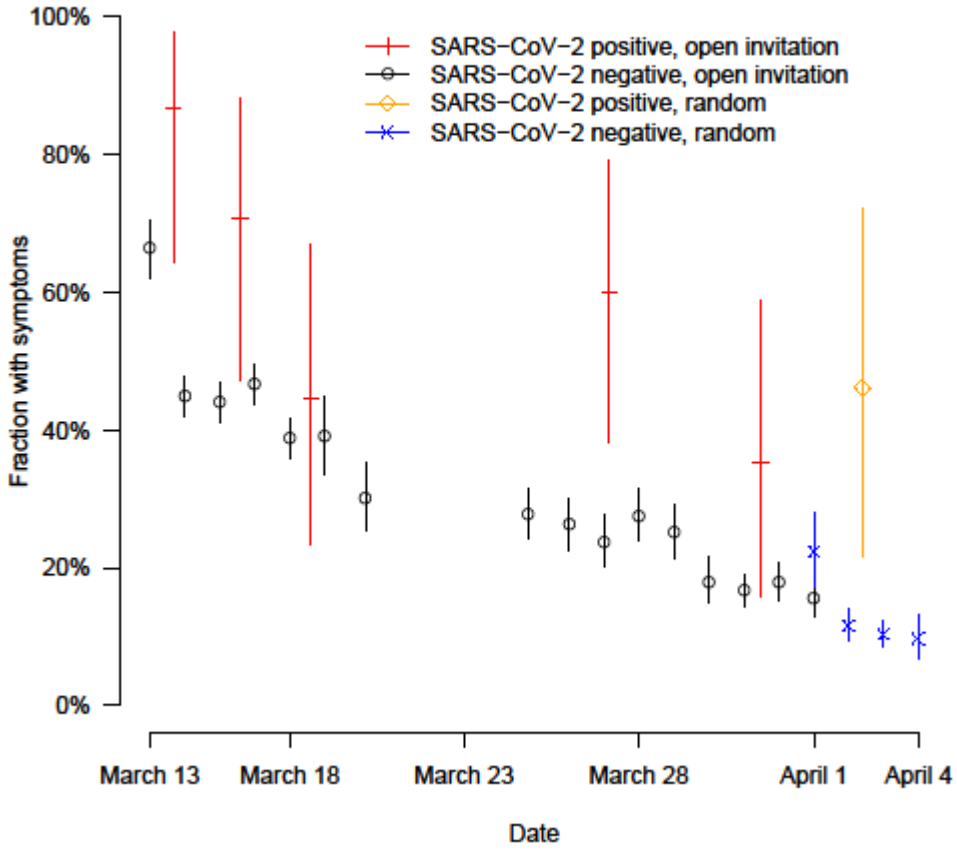


Figure S4 The fraction of participants in the population screening reporting symptoms by study date.

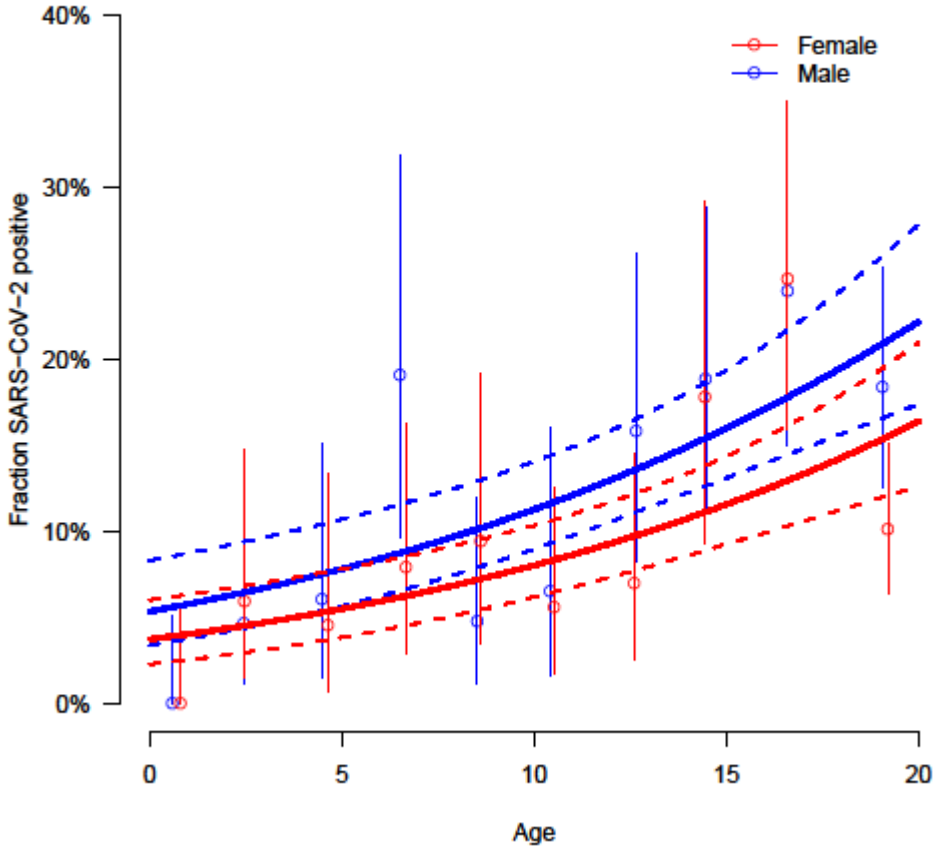


Figure S5 The fraction of individuals that tested positive before age 20 in the targeted testing stratified by age and sex. The results for males are shown in blue and females in red. Vertical bars indicate 95% confidence intervals. The solid curves indicate logistic regression fits of a model with a sex effect and an age effect. The dashed lines indicate 95% confidence intervals for the male and female logistic regression fits. The age odds ratio is 1.08 per year (95% CI: 1.05-1.12) and the male sex odds ratio is 1.45 (95% CI: 1.04-2.0).

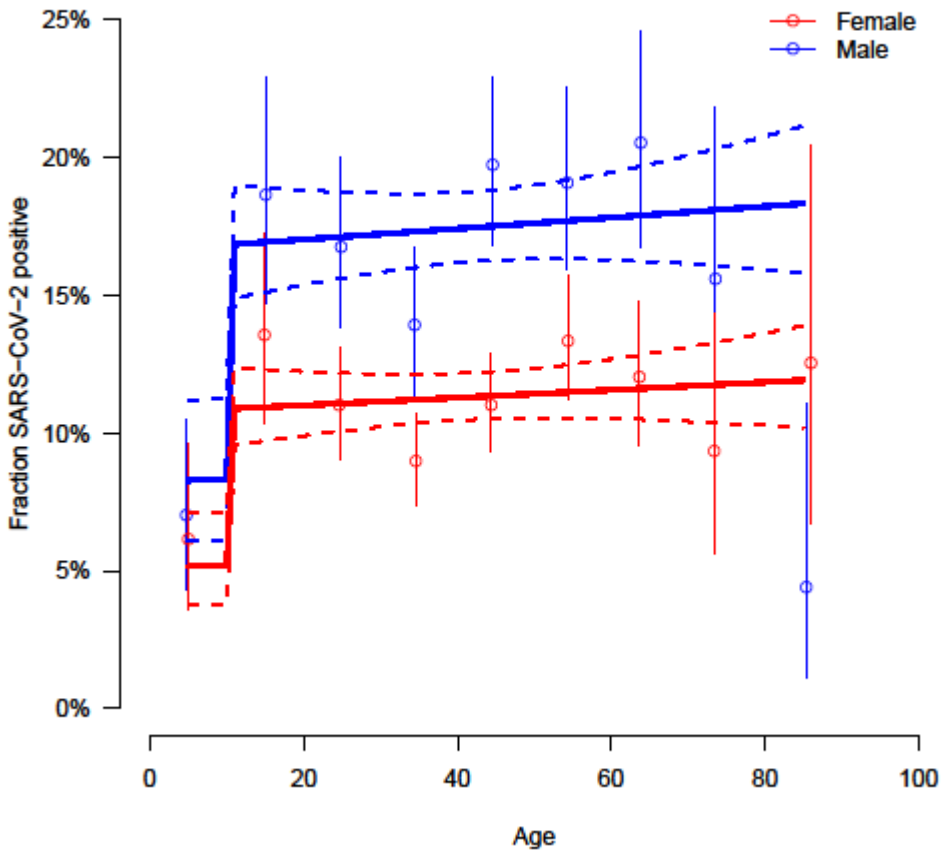


Figure S6 The fraction of individuals that tested positive in the targeted testing stratified by sex. The results for males are shown in blue and females in red. Vertical bars indicate 95% confidence intervals. The solid curves indicate logistic regression fits of a model with a sex effect, an age effect and an effect for those under 10 years of age. The dashed lines indicate 95% confidence intervals for the male and female logistic regression fits. As discussed in the main text, the fraction of positives is lower in those under 10 years of age compared to those older, but the age effect is not significant.

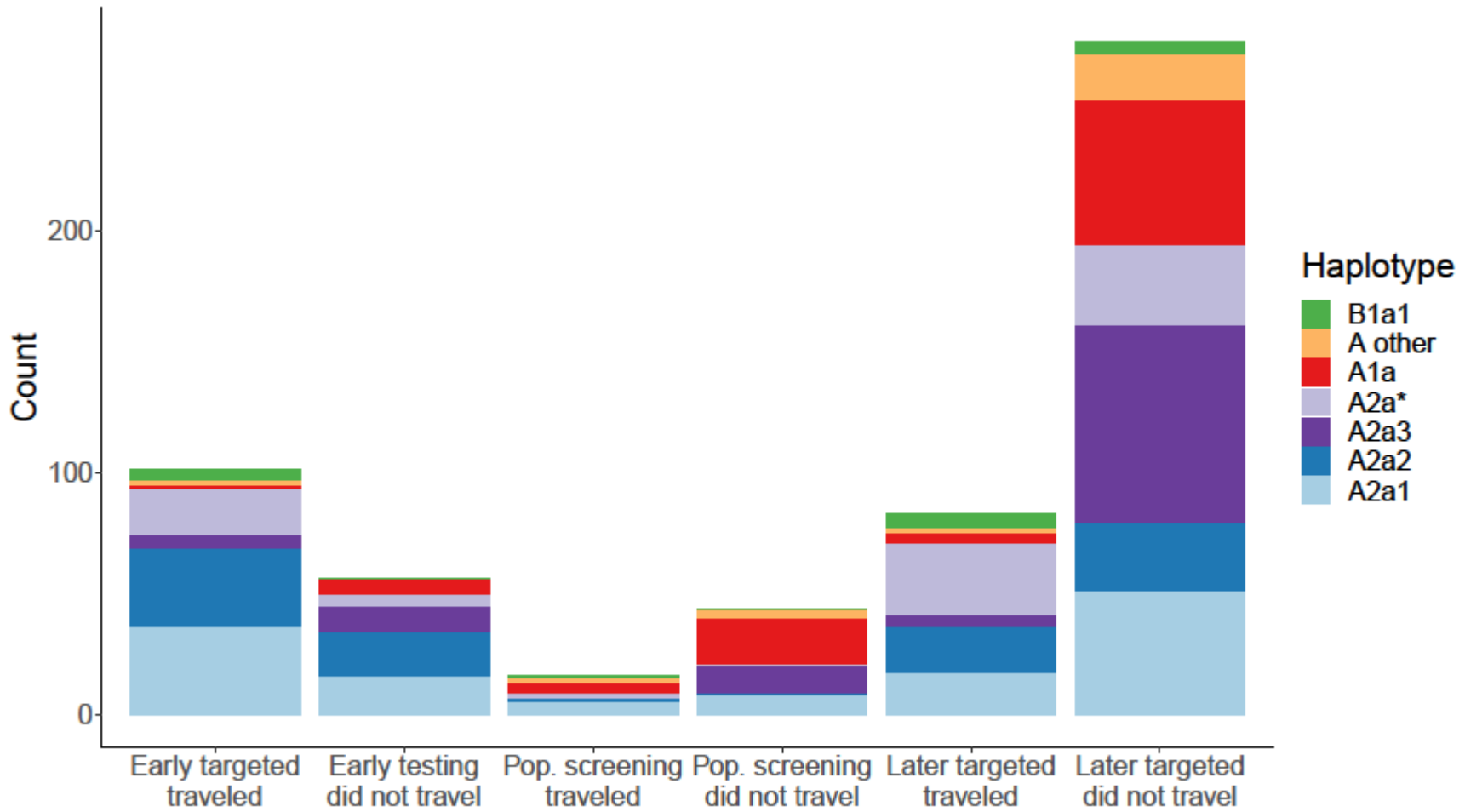


Figure S7 The distribution of SARS-CoV-2 haplotypes depending on sampling and travel status. The counts of each of the eight haplotypes seen in Iceland is shown depending on whether the positive was found through early targeted testing, population screening, or later targeted testing and whether the positive had recently traveled outside Iceland.

SUPPLEMENTARY TABLES

Table S1 Primers used for PCR amplification of viral cDNA for sequencing

name	pool	seq	length	%gc	tm
nCoV-2019_1_LEFT	nCoV-2019_1	ACCAACCAACTTTCGATCTCTTGT	24	41.67	60.69
nCoV-2019_2_RIGHT	nCoV-2019_2	TAAGGATCAGTGCCAAGCTCGT	22	50	61.74
nCoV-2019_3_LEFT	nCoV-2019_1	CGGTAATAAAGGAGCTGGTGGC	22	54.55	61.32
nCoV-2019_4_RIGHT	nCoV-2019_2	CACAAGTAGTGGCACCTTCTTTAGT	25	44	60.97
nCoV-2019_5_LEFT	nCoV-2019_1	TGGTGAAACTTCATGGCAGACG	22	50	61.39
nCoV-2019_6_RIGHT	nCoV-2019_2	TAGCGGCCTTCTGTAAAACACG	22	50	61.18
nCoV-2019_7_LEFT	nCoV-2019_1	ATCAGAGGCTGCTCGTGTTGTA	22	50	61.73
nCoV-2019_8_RIGHT	nCoV-2019_2	GCTTCAACAGCTTCACTAGTAGGT	24	45.83	60.56
nCoV-2019_9_LEFT	nCoV-2019_1	TCCCACAGAAGTGTTAACAGAGGA	24	45.83	61.18
nCoV-2019_10_RIGHT	nCoV-2019_2	TCATCTAACCAATCTTCTTCTTGCTCT	27	37.04	60.31
nCoV-2019_11_LEFT	nCoV-2019_1	GGAATTTGGTGCCACTTCTGCT	22	50	61.66
nCoV-2019_12_RIGHT	nCoV-2019_2	TTCACTCTTCATTTCCAAAAAGCTTGA	27	33.33	60.36
nCoV-2019_13_LEFT	nCoV-2019_1	TCGCACAAATGTCTACTTAGCTGT	24	41.67	60.56
nCoV-2019_14_RIGHT	nCoV-2019_2	AGTTTCCACACAGACAGGCATT	22	45.45	60.42
nCoV-2019_15_LEFT	nCoV-2019_1	ACAGTGCTTAAAAAGTGTAAAAGTGCC	27	37.04	61.32
nCoV-2019_16_RIGHT	nCoV-2019_2	CACAACTTGCGTGTGGAGGTTA	22	50	61.32
nCoV-2019_17_LEFT	nCoV-2019_1	CTTCTTTCTTTGAGAGAAGTGAGGACT	27	40.74	60.69
nCoV-2019_18_RIGHT	nCoV-2019_2	AGCTTGTTTACCACACGTACAAGG	24	45.83	61.51
nCoV-2019_19_LEFT	nCoV-2019_1	GCTGTTATGTACATGGGCACACT	23	47.83	61.18
nCoV-2019_20_RIGHT	nCoV-2019_2	ACGTGGCTTTATTAGTTGCATTGTT	25	36	60.28
nCoV-2019_21_LEFT	nCoV-2019_1	TGGCTATTGATTATAAACACTACACACCC	29	37.93	61.49
nCoV-2019_22_RIGHT	nCoV-2019_2	ACAGTATTCTTTGCTATAGTAGTCGGC	27	40.74	60.73
nCoV-2019_23_LEFT	nCoV-2019_1	ACAACTACTAACATAGTTACACGGTGT	27	37.04	60.26
nCoV-2019_24_RIGHT	nCoV-2019_2	ACATTCTAACCATAGCTGAAATCGGG	26	42.31	61.19
nCoV-2019_25_LEFT	nCoV-2019_1	GCAATTGTTTTTCAGCTATTTTGCAGT	27	33.33	60.73
nCoV-2019_26_RIGHT	nCoV-2019_2	TCCGCACTATCACCAACATCAG	22	50	60.42
nCoV-2019_27_LEFT	nCoV-2019_1	ACTACAGTCAGCTTATGTGTCAACC	25	44	60.8
nCoV-2019_28_RIGHT	nCoV-2019_2	TGTTTAGACATGACATGAACAGGTGT	26	38.46	60.91
nCoV-2019_29_LEFT	nCoV-2019_1	ACTTGTGTTCCTTTTTGTTGCTGC	24	41.67	61.39
nCoV-2019_30_RIGHT	nCoV-2019_2	ACCACTAGTAGATACACAAACACCAG	26	42.31	60.3
nCoV-2019_31_LEFT	nCoV-2019_1	TTCTGAGTACTGTAGGCACGGC	22	54.55	62.03
nCoV-2019_32_RIGHT	nCoV-2019_2	AGCACATCACTACGCAACTTTAGA	24	41.67	60.56
nCoV-2019_33_LEFT	nCoV-2019_1	ACTTTTGAAGAAGCTGCGCTGT	22	45.45	61.58
nCoV-2019_34_RIGHT	nCoV-2019_2	AGTGAAATTGGGCCTCATAGCA	22	45.45	60.03
nCoV-2019_35_LEFT	nCoV-2019_1	TGTTCGCATTCAACCAGGACAG	22	50	61.39
nCoV-2019_36_RIGHT	nCoV-2019_2	GAACAAAGACCATTGAGTACTCTGGA	26	42.31	60.74
nCoV-2019_37_LEFT	nCoV-2019_1	ACACACCACTGGTTGTTACTCAC	23	47.83	60.93
nCoV-2019_38_RIGHT	nCoV-2019_2	CACCAAGAGTCAGTCTAAAGTAGCG	25	48	61.13
nCoV-2019_39_LEFT	nCoV-2019_1	AGTATTGCCCTATTTTCTTCATAACTGGT	29	34.48	61
nCoV-2019_40_RIGHT	nCoV-2019_2	CATGGCTGCATCACGGTCAAAT	22	50	62.09
nCoV-2019_41_LEFT	nCoV-2019_1	GTTCCCTTCCATCATATGCAGCT	23	47.83	60.75
nCoV-2019_42_RIGHT	nCoV-2019_2	CCTACCTCCCTTTGTTGTGTTGT	23	47.83	60.69

nCoV-2019_43_LEFT	nCoV-2019_1	TACGACAGATGTCTTGTGCTGC	22	50	60.93
nCoV-2019_44_RIGHT	nCoV-2019_2	AACCTTTCCACATACCGCAGAC	22	50	60.87
nCoV-2019_45_LEFT	nCoV-2019_1	TACCTACAACTTGTGCTAATGACCC	25	44	60.57
nCoV-2019_46_RIGHT	nCoV-2019_2	CACGTTCACCTAAGTTGGCGTA	22	50	60.86
nCoV-2019_47_LEFT	nCoV-2019_1	AGGACTGGTATGATTTTGTAGAAAACCC	28	39.29	61.42
nCoV-2019_48_RIGHT	nCoV-2019_2	TAGATTACCAGAAGCAGCGTGC	22	50	60.74
nCoV-2019_49_LEFT	nCoV-2019_1	AGGAATTACTTGTGTATGCTGCTGA	25	40	60.57
nCoV-2019_50_RIGHT	nCoV-2019_2	TAACATGTTGTGCCAACCACCA	22	45.45	60.95
nCoV-2019_51_LEFT	nCoV-2019_1	TCAATAGCCGCCACTAGAGGAG	22	54.55	61.34
nCoV-2019_52_RIGHT	nCoV-2019_2	GTTGAGAGCAAAATTCATGAGGTCC	25	44	60.62
nCoV-2019_53_LEFT	nCoV-2019_1	AGCAAAATGTTGGACTGAGACTGA	24	41.67	60.69
nCoV-2019_54_RIGHT	nCoV-2019_2	AACCAAAAACTTGTCCATTAGCACA	25	36	60.11
nCoV-2019_55_LEFT	nCoV-2019_1	ACTCAACTTTACTTAGGAGGTATGAGCT	28	39.29	61.43
nCoV-2019_56_RIGHT	nCoV-2019_2	ACACTATGCGAGCAGAAGGGTA	22	50	61.21
nCoV-2019_57_LEFT	nCoV-2019_1	ATTCTACACTCCAGGGACCACC	22	54.55	61.16
nCoV-2019_58_RIGHT	nCoV-2019_2	CTTTTCTCCAAGCAGGGTTACGT	23	47.83	61.06
nCoV-2019_59_LEFT	nCoV-2019_1	TCACGCATGATGTTTCATCTGCA	23	43.48	61.42
nCoV-2019_60_RIGHT	nCoV-2019_2	GGTACCAACAGCTTCTCTAGTAGC	24	50	60.44
nCoV-2019_61_LEFT	nCoV-2019_1	TGTTTATCACCCGCGAAGAAGC	22	50	61.5
nCoV-2019_62_RIGHT	nCoV-2019_2	GTTGAACCTTTCTACAAGCCGC	22	50	60.35
nCoV-2019_63_LEFT	nCoV-2019_1	TGTTAAGCGTGTTGACTGGACT	22	45.45	60.16
nCoV-2019_64_RIGHT	nCoV-2019_2	AGTCTTGTAAAAGTGTTCCAGAGGT	25	40	60.1
nCoV-2019_65_LEFT	nCoV-2019_1	GCTGGCTTTAGCTTGTGGGTTT	22	50	61.92
nCoV-2019_66_RIGHT	nCoV-2019_2	TCAATTTCCATTTGACTCCTGGGT	24	41.67	60.45
nCoV-2019_67_LEFT	nCoV-2019_1	GTTGTCCAACAATTACCTGAAACTTACT	28	35.71	60.43
nCoV-2019_68_RIGHT	nCoV-2019_2	CTCCTTTATCAGAACCAGCACCA	23	47.83	60.31
nCoV-2019_69_LEFT	nCoV-2019_1	TGTCGCAAAATATACTCAACTGTGTCA	27	37.04	61.43
nCoV-2019_70_RIGHT	nCoV-2019_2	TGACCTTCTTTTAAAGACATAACAGCAG	28	35.71	60.27
nCoV-2019_71_LEFT	nCoV-2019_1	ACAAATCCAATTCAGTTGTCTTCCTATTC	29	34.48	60.54
nCoV-2019_72_RIGHT	nCoV-2019_2	ACTCTGAACTCACTTTCCATCCAAC	25	44	60.97
nCoV-2019_73_LEFT	nCoV-2019_1	CAATTTTGTAATGATCCATTTTTGGGTGT	29	31.03	60.29
nCoV-2019_74_RIGHT	nCoV-2019_2	GCAACACAGTTGCTGATTCTCTTC	24	45.83	60.85
nCoV-2019_75_LEFT	nCoV-2019_1	AGAGTCCAACCAACAGAATCTATTGT	26	38.46	60.24
nCoV-2019_76_RIGHT	nCoV-2019_2	ACACCTGTGCCTGTTAAACCAT	22	45.45	60.42
nCoV-2019_77_LEFT	nCoV-2019_1	CCAGCAACTGTTTGTGGACCTA	22	50	60.75
nCoV-2019_78_RIGHT	nCoV-2019_2	TGTGTACAAAAACTGCCATATTGCA	25	36	60.22
nCoV-2019_79_LEFT	nCoV-2019_1	GTGGTGATTCAACTGAATGCAGC	23	47.83	60.92
nCoV-2019_80_RIGHT	nCoV-2019_2	TGGAGCTAAGTTGTTTAACAAGCG	24	41.67	60.02
nCoV-2019_81_LEFT	nCoV-2019_1	GCACTTGGAAAACTTCAAGATGTGG	25	44	61.24
nCoV-2019_82_RIGHT	nCoV-2019_2	TGCCAGAGATGTCACCTAAATCAA	24	41.67	60.02
nCoV-2019_83_LEFT	nCoV-2019_1	TCCTTTGCAACCTGAATTAGACTCA	25	40	60.46
nCoV-2019_84_RIGHT	nCoV-2019_2	AGGTGTGAGTAAACTGTTACAAACAAC	27	37.04	60.36
nCoV-2019_85_LEFT	nCoV-2019_1	ACTAGCACTCTCCAAGGGTGTT	22	50	61.03
nCoV-2019_86_RIGHT	nCoV-2019_2	ACGAAAGCAAGAAAAAGAAGTACGC	25	40	61.01
nCoV-2019_87_LEFT	nCoV-2019_1	CGACTACTAGCGTGCCTTTGTA	22	50	60.16
nCoV-2019_88_RIGHT	nCoV-2019_2	TGGTCAGAATAGTGCCATGGAGT	23	47.83	61.4

nCoV-2019_89_LEFT	nCoV-2019_1	GTACGCGTTCCATGTGGTCATT	22	50	61.5
nCoV-2019_90_RIGHT	nCoV-2019_2	TGAAATGGTGAATTGCCCTCGT	22	45.45	60.82
nCoV-2019_91_LEFT	nCoV-2019_1	TCACTACCAAGAGTGTGTTAGAGGT	25	44	60.93
nCoV-2019_92_RIGHT	nCoV-2019_2	AGGTTCCTGGCAATTAATTGTAAAAGG	27	37.04	60.53
nCoV-2019_93_LEFT	nCoV-2019_1	TGAGGCTGGTTCTAAATCACCCA	23	47.83	61.59
nCoV-2019_94_RIGHT	nCoV-2019_2	TTTGGCAATGTTGTTCCTTGAGG	23	43.48	60.18
nCoV-2019_95_LEFT	nCoV-2019_1	TGAGGGAGCCTTGAATACACCA	22	50	61.1
nCoV-2019_96_RIGHT	nCoV-2019_2	TAGGCTCTGTTGGTGGGAATGT	22	50	61.36
nCoV-2019_97_LEFT	nCoV-2019_1	TGGATGACAAAGATCCAAATTTCAAAGA	28	32.14	60.22
nCoV-2019_98_RIGHT	nCoV-2019_2	TTCTCCTAAGAAGCTATTAAAATCACATGG	30	33.33	60.01

Table S2 Illumina indexes used in sequencing

Well i7 index

1. A1 CCGCGGTT
2. A2 AGTTCAGG
3. A3 TAATACAG
4. A4 ACTAAGAT
5. A5 TACCGAGG
6. A6 ATATGGAT
7. A7 ATATCTCG
8. A8 TGCGGCGT
9. A9 ACACTAAG
10. A10 CAATTAAC
11. A11 AACTGTAG
12. A12 TATCGCAC

1. B1 TTATAACC
2. B2 GACCTGAA
3. B3 CGGCGTGA
4. B4 GTCGGAGC
5. B5 CGTTAGAA
6. B6 GCGCAAGC
7. B7 GCGCTCTA
8. B8 CATAATAC
9. B9 GTGTCGGA
10. B10 TGGCCGGT
11. B11 GGTCACGA
12. B12 CGCTATGT

1. C1 GGACTTGG
2. C2 TCTCTACT
3. C3 ATGTAAGT
4. C4 CTTGGTAT
5. C5 AGCCTCAT
6. C6 AAGATACT
7. C7 AACAGGTT
8. C8 GATCTATC
9. C9 TTCCTGTT
10. C10 AGTACTCC
11. C11 CTGCTTCC
12. C12 GTATGTTC

1. D1 AAGTCCAA
2. D2 CTCTCGTC
3. D3 GCACGGAC
4. D4 TCCAACGC

i5 index

AGCGCTAG CCAACAGA ATATTCAC AACCGCGG CCTGAACT CTGTATTA ATCTTAGT CCTCGGTA
ATCCATAT CGAGATAT ACGCCGCA CTTAGTGT GATATCGA TTGGTGAG GCGCCTGT GGTTATAA
TTCAGGTC TCACGCCG GCTCCGAC TTCTAACG GCTTGCGC TAGAGCGC GTATTATG TCCGACAC
CGCAGACG CGCGGTTC ACTCTATG CCAAGTCC AGTAGAGA ACTTACAT ATACCAAG ATGAGGCT
AGTATCTT AACCTGTT GATAGATC AACAGGAA TATGAGTA TATAACCT GTCTCGCA TTGGACTT

5. D5 GATTCTGC
6. D6 GGAGCGTC
7. D7 GGTGAACC
8. D8 AGCTCGCT
9. D9 CCTTCACC
10. D10 GACGTCTT
11. D11 TCATCCTT
12. D12 ACGCACCT

1. E1 ATCCACTG
2. E2 CCAAGTCT
3. E3 GGTACCTT
4. E4 CCGTGAAG
5. E5 TCGTAGTG
6. E6 ATGGCATG
7. E7 CAACAATG
8. E8 CGGAACTG
9. E9 GCCACAGG
10. E10 TGCGAGAC
11. E11 AGGTTATA
12. E12 TACTCATA

1. F1 GCTTGTCA
2. F2 TTGGACTC
3. F3 AACGTTCC
4. F4 TTACAGGA
5. F5 CTACGACA
6. F6 GCAATGCA
7. F7 TGGTGGCA
8. F8 TAAGGTCA
9. F9 ATTGTGAA
10. F10 CATAGAGT
11. F11 GAACCGCG
12. F12 CGTCTGCG

1. G1 CAAGCTAG
2. G2 GGCTTAAG
3. G3 GCAGAATT
4. G4 GGCATTCT
5. G5 TAAGTGGT
6. G6 GTTCCAAT

7. G7 AGGCAGAG
8. G8 TTGCCTAG
9. G9 ACTCGTGT
10. G10 ACAGGCGC
11. G11 CTCACCAA

GACGAGAG GTCCGTGC GCGTTGGA GCAGAATC GACGCTCC GGTTCACC AGCGAGCT GGTGAAGG AGGTGCGT AAGGATGA AAGACGTC CAGTGGAT AGACTTGG AAGGTACC CTTCACGG CACTACGA CATGCCAT CATTGTTG CAGTTCCG CCTGTGGC GAACATAC GGAAGCAG GGAGTACT TGACAAGC GAGTCCAA GGAACGTT TCCTGTAA TGTCGTAG TGCATTGC TGCCACCA TGACCTTA TTCACAAT ACATAGCG TCGTGACC ACCGGCCA CTAGCTTG CTTAAGCC AATTCTGC AGAATGCC ACCACTTA ATTGGAAC CTCTGCCT CTAGGCAA

G12 TCGATATC

1. H1 TGGATCGA
2. H2 AATCCGGA
3. H3 ATGAGGCC
4. H4 AATGCCTC
5. H5 CGGACAAC
6. H6 ACCTTGGC
7. H7 GAATGAGA
8. H8 CCATTCGA
9. H9 GTCTACAC
10. H10 GTGAATAT
11. H11 TCTGTTGG
12. H12 CTAGCGCT

ACACGAGT GTGCGATA CTACAGTT GTTAATTG TCGATCCA TCCGGATT GGCCTCAT GAGGCATT GTTGTCCG GCCAAGGT TCTCATTC TCGAATGG GTGTAGAC

Table S3 Sequence variants used to define SARS-CoV-2 clades

Clade	Pos	Ref	Alt
B	28144	T	C
B1	18060	C	T
B1a	17858	A	G
B1a1	17747	C	T
B1a1a	24694	A	T
B1a1a1	9445	T	C
B1a1a1a	17531	T	C
B1a1a1a	18756	G	T
B1a1a1b	29140	G	T
B4	28878	G	A
B4	29742	G	A
B2	29095	C	T
A	20229	C	T
A	13064	C	T
A	18483	T	C
A	8017	A	G
A1a	11083	G	T
A1a	26144	G	T
A1a1	14805	C	T
A1a1a	17247	T	C
A1a1a1	5142	C	T
A1a1a2	1321	A	C
A1a1a3	3034	T	C
A1a1a3	16054	C	T
A1a1a3	17859	T	C
A1a1a3	29751	G	C
A1a1a4	1515	A	G
A1a1a5	7479	A	G
A1a1b	2558	C	T
A1a2	7876	T	A
A3	1397	G	A
A3	11083	G	T
A3	28688	T	C
A3	29742	G	T
A6	514	T	C
A7	9924	C	T
A8	1440	G	A
A8	2891	G	A
A8a	28851	G	T
A8b	4140	A	G
A8c	27661	C	T
A9	1604	ATGA	A
A9	20270	C	T
A10	1218	C	T

A10	27806	G	T
A10	29711	G	T
A2	241	C	T
A2	3037	C	T
A2	23403	A	G
A2a	14408	C	T
A2a1	28881	G	A
A2a1	28882	G	A
A2a1	28883	G	C
A2a1a	27046	C	T
A2a1a1	25958	A	G
A2a1a2	28344	C	A
A2a1a3	23086	C	T
A2a1b	10097	G	A
A2a1b	23731	C	T
A2a1c	19839	T	C
A2a1d	27430	G	A
A2a2	25563	G	T
A2a2a	1059	C	T
A2a3	20268	A	G
A2a3a	10323	A	G
A2a4	2455	C	T
A2a4	10450	C	T
A2a5	26530	A	G
A2a6	24862	A	G
A2a7	25429	G	T
A2a8	15324	C	T
A2a9	187	A	G
A2a10	25350	C	T
A2a11	20275	G	A
A2a12	24077	G	T
A2a13	28836	C	T

The clade nomenclature at ncov-NextStrain (https://academic.oup.com/bioinformatics/article/34/23/4121/5001388) was used as a basis for our clade definitions and we added clades in accordance with the phylogenetic tree at NextStrain (at 2020-04-04).

Table S4 GISAID sequences 23.03.2020: List of contributing labs. See excel document.

REFERENCES

1. Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020;579(7798):270–3.
2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–60.
3. Broad Institute [Internet]. Available from: http://broadinstitute.github.io/picard/;
4. Eggertsson HP, Jonsson H, Kristmundsdottir S, et al. Graphtyper enables population-scale genotyping using pangenome graphs. Nat Genet 2017;49(11):1654–60.
5. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics [Internet] 2010 [cited 2020 Apr 5];26(16):2069–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20562413
6. Kent WJ, Sugnet CW, Furey TS, et al. The Human Genome Browser at UCSC. Genome Res 2002;12(6):996–1006.
7. Guðbjartsson H, Georgsson GF, Guðjónsson SA, et al. GORpipe: a query tool for working with sequence data based on a Genomic Ordered Relational (GOR) architecture. Bioinformatics [Internet] 2016 [cited 2020 Apr 5];32(20):3081–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27339714